無制約最適化問題に対するハイブリッド型共役勾配法の大域的収束性について (数値解析と新しい情報技術)

数理解析研究所講究録 1362 巻 2004 年 95-103

無制約最適化問題に対するハイブリッド型共役勾配法の大域的収束性について

東京理科大学大学院　富塚 博崇 (Hirotaka Tomiduka)
Tokyo University of Science
東京理科大学　矢部 博 (Hiroshi Yabe)
Department of Mathematical Information Science
Tokyo University of Science

1 はじめに

無制約最小化問題　minimize $f(x)$, $x \in R^n$ を考える。ただし、$f: R^n \to R$ は滑らかな関数で、その勾配ベクトル $\nabla f(x)$ は利用できるものとし、それを g とおく。共役勾配法のアルゴリズムは以下のとおりである。

[共役勾配法のアルゴリズム (CG)]

Step0. 初期点 x_0 を与える。初期探索方向を $d_0 = -g_0$、$k := 0$ とおく。

Step1. 収束判定を行う。

Step2. 直線探索によりステップ幅 α_k を計算して、点 x を　$x_{k+1} = x_k + \alpha_k d_k$ と更新する。

Step3. β_{k+1} を計算して、探索方向 d を　$d_{k+1} = -g_{k+1} + \beta_{k+1} d_k$ と更新する

Step4. $k \leftarrow k+1$ として、Step1 へ行く。

ここで、ステップ幅 α_k を直線探索により計算する際には次のような Wolfe 条件を課す。(ただし、$0 < \delta < \sigma < 1$)

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \tag{1}$$

$$g(x_k + \alpha_k d_k)^T d_k > \sigma g_k^T d_k \tag{2}$$

また、後に述べる定理のために strong Wolfe 条件も紹介しておく。

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \tag{3}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k| \tag{4}$$

共役勾配法は上記アルゴリズムの通り、行列を保存する必要がないため大規模な問題を解くために有効な方法である。また、線型方程式を解く共役勾配法（凸二次関数最小化問題を解く共役勾配法）と一般の無制約最小化問題を解く共役勾配法を区別するために、前者は線形共役勾配法、後者は非線形共役勾配法と呼ばれている。非線形共役勾配法は、パラメータ β_{k+1} の選び方によっていろいろな種類が考えられ、よく知られている β として

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \quad \beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$$

などがあり、それぞれ Fletcher-Reeves(FR),Polak-Ribiére-Polyak(PRP), Hestenes-Stiefel(HS) によって提案された。

また、探索方向 d_{k+1} は降下方向であることが望ましいが共役勾配法は必ずしも降下方向が保証されていないため、現在では大域的収束性の解析において、２通りのアプローチが考えられている。一

つは降下方向を仮定した上で収束性を示すアプローチで、$\beta_k^{PRP}, \beta_k^{HS}$ などが含まれる。もう一つは降下方向となるように β_{k+1} を定めるアプローチで、β_k^{FR} などが含まれている。

そこで本研究では、前者のアプローチで Yabe and Takano [6] が提案した β_{k+1} と、後者のアプローチで Yabe and Sakaiwa [5] が提案した β_{k+1} を組み合わせた新しい共役勾配法を提案し、その大域的収束性を示す。

この、２種類の方法を紹介する際に必要となる条件を簡単に述べる。

まず、探索方向が満たす共役性条件について説明する。正定値対称行列 $A \in R^{n\times n}$ をヘッセ行列にもつ凸二次関数最小化問題に対して探索方向が

$$d_i^T A d_j = 0\ ,\ (i \neq j)$$

を満たすとき、探索方向は行列 A に関して共役であるという。一方、一般の非線形関数最小化問題に対しては平均値の定理を用いることにより

$$d_{k+1}^T g_{k+1} = d_{k+1}^T g_k + \alpha_k d_{k+1}^T \nabla^2 f(x_k + \tau\alpha_k d_k) d_k,\ \tau \in (0,1)$$

と表すことができる。これを変形し、$y_k = g_{k+1} - g_k$ とおくと、

$$\alpha_k^{-1} d_{k+1}^T y_k = d_{k+1}^T \nabla^2 f(x_k + \tau\alpha_k d_k) d_k$$

となる。そこで、線形共役勾配法の場合の前述の共役性条件 $d_{k+1}^T A d_k = 0$ に対応させれば、非線形共役勾配法に対する共役性条件は

$$d_{k+1}^T y_k = 0$$

となり、行列を使わない形で表すことができる。

次に、共役勾配法のより速い収束を期待して、ヘッセ行列の情報を取り入れることを考える。具体的には、準ニュートン法においてヘッセ行列の情報を取り入れるためのセカント条件を用いる。準ニュートン法の探索方向は $d_{k+1} = -H_{k+1} g_{k+1}$ と更新される。ただし、H_{k+1} はヘッセ行列の逆行列を近似する正定値対称行列である。セカント条件とは、g_k のテイラー展開

$$g_k = g_{k+1} + \nabla^2 f(x_{k+1})(x_k - x_{k+1}) + \cdots$$

を $x_k - x_{k+1}$ の項で打ち切って $s_k = x_{k+1} - x_k$ とおけば $s_k \approx \nabla^2 f(x_{k+1})^{-1} y_k$ となるので、近似行列 H_{k+1} を用いて $H_{k+1} y_k = s_k$ と表される条件のことである。さらに、準ニュートン法の分野においてはセカント条件の改良が試みられており、Zhang ら ([7] , [8]) はその拡張として修正セカント条件を提案した。具体的には、f_k と $g_k^T s_k$ をそれぞれ３次の項までテイラー展開すると

$$\begin{aligned} f_k &= f_{k+1} - g_{k+1}^T s_k + \frac{1}{2!} s_k^T \nabla^2 f_{k+1} s_k - \frac{1}{3!} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4) \\ g_k^T s_k &= g_{k+1}^T s_k - s_k^T \nabla^2 f_{k+1} s_k + \frac{1}{2!} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4) \end{aligned}$$

となり、両式からテンソル項 T_{k+1} を消去すると

$$s_k^T \nabla^2 f_{k+1} s_k = s_k^T y_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k + O(\|s_k\|^4)$$

となる。ここで、近似行列 H_{k+1} を用いると修正セカント条件は、

$$H_{k+1}\widehat{y}_k = s_k\ ,\ \begin{cases} \widehat{y}_k = y_k + \dfrac{\theta_k}{s_k^T u_k} u_k & (u_k \text{は } s_k^T u_k \neq 0 \text{ となる任意のベクトル}) \\ \theta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k \end{cases} \tag{5}$$

と表される。さらに、Zhang らによって曲率の近似の精度が上がることも示されている。

2 Yabe and Sakaiwa Method

d_k が降下方向、すなわち $g_k^T d_k < 0$ であると仮定し d_{k+1} が降下方向となるような β_{k+1} を求める。つまり、d の更新式に左から g_{k+1}^T をかけて

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k < 0 \tag{6}$$

をみたす β_{k+1} を求める。ここで、パラメータ $\tau_{k+1} > 0$ を導入し $\beta_{k+1} = \|g_{k+1}\|^2/\tau_{k+1}$ とすると、式 (6) は、$\tau_{k+1} > g_{k+1}^T d_k$ と置き換わる。したがって、$\tau_{k+1} > 0$ を考慮すると、$\tau_{k+1} > \max\{g_{k+1}^T d_k, 0\}$ ならば降下方向となる。特に $\tau_{k+1} = d_k^T y_k$ とすると、Dai and Yuan [2] が提案した β_{k+1} となる。このとき、直線探索に Wolfe の条件を課せば $d_k^T y_k > 0$ が保証される。

Yabe and Sakaiwa [5] は、DY 法にパラメータ付き修正セカント条件を導入し、さらに式を簡単にするために $s_k = \alpha_k d_k$ を用いて

$$\tau_{k+1}^{YS} = d_k^T y_k + \frac{\lambda_k}{\alpha_k}\max\{\theta_k, 0\}, \quad \lambda_k \geq 0$$

とおくことにより、

$$\beta_{k+1}^{YS} = \frac{\|g_{k+1}\|^2}{\tau_k^{YS}} = \frac{\|g_{k+1}\|^2}{d_k^T y_k + \frac{\lambda_k}{\alpha_k}\max\{\theta_k, 0\}} \tag{7}$$

を提案した。ここで、$\tau_{k+1}^{YS} \geq d_k^T y_k$ なので、$\tau_{k+1}^{YS} > 0$ が保証される。

この β_{k+1}^{YS} を用いた共役勾配法の大域的収束性を紹介するために、次の 2 つの条件を仮定する。

Assumption 2.1

(A1) 準位集合 $L = \{x \in R | f(x) \leq f(x_0)\ \}$ は有界である。

(A2) L の近傍 N で、f は連続微分可能で g はリプシッツ連続である。

この Assumption のもとで、一般的にアルゴリズム (CG) に対して次の Lemma が成り立つ。

Lemma 2.2

Assumption 2.1 を仮定し、d_k は降下方向であると仮定する。さらに α_k は *Wolfe* の条件 *(1)-(2)* を満たすものとする。このとき、アルゴリズム *(CG)* によって生成される点列と探索方向について

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \tag{8}$$

が成り立つ。

式 (8) を Zoutendijk 条件と呼ぶ。この Lemma 2.2 を用いると、次の定理が示される。

Theorem 2.3

Assumption 2.1 を仮定する。α_k は *Wolfe* の直線探索によって得られるものとする。このとき、β_{k+1}^{YS} を用いた共役勾配法 *(CG)* は常に降下方向を生成し

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

が成り立つ。

さらに、次の定理も証明されている。

Theorem 2.4
Assumption 2.1 を仮定する。$\{x_k\}$ は アルゴリズム *(CG)* で生成されるものとする。$\tau_{k+1}^{YS} \geq d_k^T y_k$ として、α_k は *strong Wolfe* 条件 *(3)-(4)*、*(ただし* $0 < \sigma < \frac{1}{2}$*)* を満たすものとすると、十分な降下条件

$$g_k^T d_k \leq -c\|g_k\|^2,\ for\ all\ k \geq 1,$$

(ただし $c > 0$) が成り立つ。

3 Yabe and Takano Method

非線形共役勾配法における従来の共役性条件は $d_{k+1}^T y_k = 0$ であるが、Perry [4] は準ニュートン法の探索方向が $d_{k+1} = -H_{k+1}^T g_{k+1}$ であること、およびセカント条件を考慮し、

$$d_{k+1}^T y_k = -(H_{k+1} g_{k+1})^T y_k = -g_{k+1}^T (H_{k+1} y_k) = -g_{k+1}^T s_k$$

と式変形した。さらに Dai and Liao [1] はパラメータ $t \geq 0$ を導入することにより新たな共役性条件として

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k$$

を提案した。

Yabe and Takano [6] は、Dai and Liao の考え方に基づき、修正セカント条件を考慮した方法を提案した。具体的には、修正セカント条件にパラメータ $\rho \geq 0$ を導入し、

$$z_k = y_k + \rho \frac{\theta_k}{s_k^T u_k} u_k \tag{9}$$

と定義したときの条件 $d_{k+1}^T z_k = -t g_{k+1}^T s_k$ $(t \geq 0)$ を用いた。そして、この条件を満たす探索方向を生成するために、この条件に d の更新式を代入して整理し

$$\beta_{k+1}^{YT} = \frac{g_{k+1}^T (z_k - t s_k)}{d_k^T z_k}$$

という β_{k+1} を提案した。さらに、大域的収束性の観点から

$$\beta_{k+1}^{YT+} = \max\left\{\frac{g_{k+1}^T z_k}{d_k^T z_k}, 0\right\} - t\frac{g_{k+1}^T s_k}{d_k^T z_k} \tag{10}$$

という第一項を補正した β_{k+1} を提案した。

この β_{k+1}^{YT+} を用いた共役勾配法 (CG) の大域的収束性について以下の定理が得られる。

Theorem 3.1
Assumption 2.1 を仮定する。*(10)* を用いた共役勾配法 *(CG)* を考え、d_k と u_k は以下の式を満たすものとする。

$$\begin{aligned} g_k^T d_k &\leq -c\|g_k\|^2, \\ |s_k^T u_k| &\geq m\|s_k\|\|u_k\| \end{aligned}$$

ただし、c と m は正の定数である。また、α_k は、*strong Wolfe* の条件 *(3)-(4)* を満たすものとする。このとき、$0 \leq \rho < \frac{1-\sigma}{3(1+\sigma-2\delta)}$ ならば、

$$\liminf_{k\to\infty} \|g_k\| = 0$$

となる。

4 新しいβの提案とその大域的収束性

2節で述べたYabe and Sakaiwa 法は降下方向の保証があり、他方、3節で述べたYabe and Takano 法は降下方向の保証はないもののセカント条件（曲率）の情報をより効果的に利用している。いずれの方法でも良い数値実験結果が得られているので、この2つをうまく組み合わせることによって、降下方向を生成し、かつ、曲率の情報をある程度取り込んだ共役勾配法を作ることが考えられる。

まず、$\beta_{k+1} \geq 0$ を仮定する。(6) の条件を

$$\begin{aligned} -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k &= -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k - \beta_{k+1} g_k^T d_k + \beta_{k+1} g_k^T d_k \\ &= -\|g_{k+1}\|^2 + \beta_{k+1} d_k^T y_k + \beta_{k+1} g_k^T d_k \\ &< 0 \end{aligned}$$

と変形し $g_k^T d_k < 0$ を考慮すると、降下方向を生成する条件 (6) は

$$\|g_{k+1}\|^2 \geq \beta_{k+1} d_k^T y_k \tag{11}$$

と置き換わる。今後は (11) を満たす β_{k+1} を考える。そこで、前述の2つのパラメータ $\beta_{k+1}^{YT+}, \beta_{k+1}^{YS}$ の凸結合を作り、β_{k+1} として

$$\begin{aligned} \beta_{k+1}^{New} &= \phi_k \beta_{k+1}^{YT+} + (1-\phi_k)\beta_{k+1}^{YS} \\ &= \phi_k \left\{ \max\left\{ \frac{g_{k+1}^T z_k}{d_k^T z_k}, 0 \right\} - t\frac{g_{k+1}^T s_k}{d_k^T z_k} \right\} + (1-\phi_k)\frac{\|g_{k+1}\|^2}{d_k^T y_k + \frac{\lambda_k}{\alpha_k}\max\{\theta_k, 0\}} \end{aligned} \tag{12}$$

を提案する。ただし、$0 < \phi_k < 1$ であり、パラメータ t は $\beta_{k+1}^{YT+} \geq 0$ を満たすように

$$\begin{cases} \text{(i)} \quad \frac{g_{k+1}^T s_k}{d_k^T z_k} \leq 0 \text{のとき} \quad 0 \leq t \\ \text{(ii)} \quad \frac{g_{k+1}^T s_k}{d_k^T z_k} > 0 \text{のとき} \quad 0 \leq t \leq \frac{d_k^T z_k}{g_{k+1}^T s_k} \max\left\{ \frac{g_{k+1}^T z_k}{d_k^T z_k}, 0 \right\} \end{cases} \tag{13}$$

と定めるものとする。このとき、探索方向 d_{k+1}^{New} は Yabe and Sakaiwa の探索方向 d_{k+1}^{YS} と Yabe and Takano の探索方向 d_{k+1}^{YT+} を用いて

$$d_{k+1}^{New} = \phi_k d_{k+1}^{YT+} + (1-\phi_k) d_{k+1}^{YS} \tag{14}$$

と表せる。次に、(12) の β_{k+1}^{New} が降下方向の条件 (11) を満たすようにパラメータ ϕ_k を決める。すなわち、(12) を (11) に代入すると

$$\begin{aligned} \|g_{k+1}\|^2 &\geq \beta_{k+1} d_k^T y_k \\ &= \left\{ \phi_k \left(\max\left\{ \frac{g_{k+1}^T z_k}{d_k^T z_k}, 0 \right\} - t\frac{g_{k+1}^T s_k}{d_k^T z_k} \right) + (1-\phi_k)\frac{\|g_{k+1}\|^2}{\tau_{k+1}^{YS}} \right\} d_k^T y_k \\ &= \phi_k \left\{ \max\left\{ \frac{g_{k+1}^T z_k}{d_k^T z_k}, 0 \right\} - t\frac{g_{k+1}^T s_k}{d_k^T z_k} - \frac{\|g_{k+1}\|^2}{\tau_{k+1}^{YS}} \right\} d_k^T y_k + \frac{\|g_{k+1}\|^2}{\tau_{k+1}^{YS}} d_k^T y_k \end{aligned} \tag{15}$$

となり、この式を整理すると

$$\frac{\tau_{k+1}^{YS} - d_k^T y_k}{\tau_{k+1}^{YS}} \|g_{k+1}\|^2 \geq \phi_k \eta_k d_k^T y_k, \tag{16}$$

$$\eta_k \equiv \max\left\{ \frac{g_{k+1}^T z_k}{d_k^T z_k}, 0 \right\} - t\frac{g_{k+1}^T s_k}{d_k^T z_k} - \frac{\|g_{k+1}\|^2}{\tau_{k+1}^{YS}}$$

が得られる。ここで、$\tau_{k+1}^{YS} \geq d_k^T y_k$ より左辺の値は非負である。したがって、式 (16) を満たすような ϕ_k を選べば探索方向 (14) が降下方向になることがわかる。この事実に基づいて以下の定理を得る。

Theorem 4.1
η_k の値に応じて、ϕ_k を次のように選ぶ。

$$\begin{cases} (\mathrm{i}) & \eta_k \leq 0 \text{ のとき } \quad 0 \leq \phi_k \leq 1 \\ (\mathrm{ii}) & \eta_k > 0 \text{ のとき } \quad 0 \leq \phi_k \leq \min\left\{ \frac{1}{\eta_k d_k^T y_k} \frac{\tau_{k+1}^{YS} - d_k^T y_k}{\tau_{k+1}^{YS}} \|g_{k+1}\|^2, 1 \right\} \end{cases} \tag{17}$$

このとき、我々の提案する探索方向 *(14)* は降下方向となる。

このとき、β_{k+1}^{New} を用いた共役勾配法の大域的収束性について次の定理を得る。

Theorem 4.2
Assumption 2.1 を仮定する。β_{k+1}^{New} は *(12)* と *(17)* で定義され、α_k は *Wolfe* の直線探索によって得られるものとする。このとき、β_{k+1}^{New} を用いた共役勾配法 *(CG)* は常に降下方向を生成し

$$\liminf_{k \to \infty} \|g_k\| = 0.$$

が成り立つ。

さらに、「十分な降下条件」についての次の定理も証明することができる。

Theorem 4.3
Assumption 2.1 を仮定する。$\{x_k\}$ は 上述のアルゴリズムで生成されるものとする。β_{k+1} として、$\beta_{k+1} \geq 0, \|g_{k+1}\|^2 \geq \beta_{k+1} d_k^T y_k$ を満たすものを選び、α_k は *strong Wolfe* 条件 *(3)-(4)* を満たすものとすると、十分な降下条件

$$g_k^T d_k \leq -c\|g_k\|^2, \; for \; all \; k \geq 1,$$

(ただし $c \geq 0$) が成り立つ。

5 数値実験結果

テスト関数として用いたのは、Moré らの論文 [3] に掲載されている拡張 Rosenbrock 関数

$$f(x) = \sum_{i=1}^{n/2} 100(x_{2i} - x_{2i-1}^2)^2 + \sum_{i=1}^{n/2} (1 - x_{2i-1})^2$$

で、ここでは次元は $n = 1{,}000$ とする。また、初期点は $(-1.2, 1, -1.2, 1, \cdots, -1.2, 1)$ とし、パラメータ付きセカント条件 (9) に含まれる任意のベクトル u_k として s_k, y_k, g_{k+1}, g_k を用いる。収束判定条件は、$\|\nabla f(x_k)\|_\infty < 10^{-5}$ とする。さらに、直線探索を行う際には簡単のために (1) だけを用いており、パラメータは $\delta = 0.01$ を用いた。

提案した β_{k+1} には4つのパラメータが含まれているので、各パラメータの決め方を記す。

式 (17) で与えられた凸結合パラメータ ϕ_k は、取り得る限りは0.5とし、無理な場合は $\widehat{\phi_k}$ または0を用いる。(ただし、$\widehat{\phi_k} = \frac{1}{\eta_k d_k^T y_k} \frac{\tau_{k+1}^{YS} - d_k^T y_k}{\tau_{k+1}^{YS}} \|g_{k+1}\|^2$ である。) 理論的には負になるような場合は考えにくいが、実験では Wolfe の条件を用いていないため、$\widehat{\phi_k} < 0$ となることもある。この場合には $\phi_k = 0$ とした。共役性条件に用いられる t は (13) を満たすような値を選ぶが、実験においてはまず定数を与えておき条件が満たされないときには0とした。実際に0を取ることもあったが、パラメータの選び方によって回数に大きな差がある。修正セカント条件に含まれるパラメータ ρ と λ はそれぞれ $0 \leq \rho \leq 1, 0 \leq \lambda \leq 1$ の範囲で動かした。

以下に示す表は、パラメータ λ, ρ, t を 0.1 刻みで動かしていったときの良い場合と悪い場合の例である。具体的に、Table 1 と Table 2 の読み方は、左からパラメータ λ, ρ, t の値、そして、そのときの 反復回数/関数評価回数、さらに、ϕ_k としてどのような値が何回とられたかを示している。Table 3 は既存の方法との反復回数・関数評価回数を比較したもので、条件の欄に書いてある値は最も良い結果を与えたパラメータの値である。

まず、$u_k = s_k$ としたときの良い数値結果を与えたパラメータの例を示す。

Table 1：良い例

λ	ρ	t		$\phi_k = 0.5$	$\phi_k = \widehat{\phi_k}$	$\phi_k = 0$
0	0.4	0.4	26/91	11	0	15
0.1	0.2	0.5	26/90	7	9	10
0.1	0.9	0.4	23/79	12	6	5
0.1	0.9	0.7	21/74	9	6	6
0.2	0.8	0.2	25/86	10	7	8
0.2	1	0.9	26/86	14	5	7
0.3	0.9	1	25/89	15	6	4
0.7	0.7	0.5	24/79	13	5	6
0.7	0.8	0.5	26/87	12	7	7
0.8	0.4	0.5	24/84	10	7	7

続いて、$u_k = s_k$ としたときの悪い数値結果を与えたパラメータの例を示す。

Table 2：悪い例

λ	ρ	t		$\phi_k = 0.5$	$\phi_k = \widehat{\phi_k}$	$\phi_k = 0$
0.1	0.3	0.5	304/1201	12	146	146
0.1	0.1	0.9	273/1081	9	130	134
0.3	0.2	0.2	291/1152	7	142	142
0.3	1	0.7	280/1107	12	133	135
0.8	0.3	0.6	272/1074	8	132	135
0.9	0.8	0.8	266/1043	13	125	128
1	0.3	0.4	276/1087	8	134	134
1	0.7	0.1	285/1125	15	135	135
1	0.9	1	285/1121	17	134	134
1	1	0.4	300/1181	17	140	143

さらに、既存の方法と比較したのが以下の表である。

Table 3 : 既存の方法との比較

β		条件	降下方向
FR	85/358		生成
HS	34/220		仮定
PRP	35/189		仮定
DY	83/370		生成
$DL+$	29/94	$t=1$	仮定
YS	43/146	$\lambda=0.3$	生成
$YT+$	20/61	$\rho=1, t=0.3$	仮定
$Hybrid$	21/74	$\lambda=0.1, \rho=0.9, t=0.7$	生成

Table 3 中に登場する FR,HS,PRP はそれぞれ $\beta^{FR}, \beta^{HS}, \beta^{PRP}$ を表しており、DY,DL+ はそれぞれ Dai and Yuan, Dai and Liao が提案した β で、次式で与えられる。

$$\beta_{k+1}^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}, \quad \beta_{k+1}^{DL+} = \max\left\{\frac{g_{k+1}^T y_k}{d_k^T y_k}, 0\right\} - t\frac{g_{k+1}^T s_k}{d_k^T y_k}.$$

また、YS,YT+ はそれぞれ β^{YS}, β^{YT+} を意味している。

6 考察

Table 3 からわかるように、降下方向を仮定するアプローチの方が反復回数が少なくなる傾向があることが知られていた。しかし、今回の実験結果 Table 1 を見ると良いパラメータを見つけることができれば降下方向を生成するアプローチでも良い結果が得られることがわかった。ただし、Table 2 を見てわかる通り、パラメータの選び方によっては悪くなってしまうこともある。

ただ一つの例だけでは傾向をつかむことはできないので、同じ 1000 次元の拡張 Rosenbrock 関数に対して $u_k = y_k$ を用いた時の結果を同じように 2 つの表にまとめると、

Table 4 : 良い例

λ	ρ	t		$\phi_k = 0.5$	$\phi_k = \widehat{\phi_k}$	$\phi_k = 0$
0.1	0.2	0.5	26/88	10	8	8
0.2	0.2	0.2	24/85	10	9	10
0.2	0.2	0.4	28/99	13	7	8
0.2	0.2	0.6	28/97	13	7	8
0.2	0.2	1.0	25/88	11	6	8
0.2	0.4	0.9	28/97	10	9	9
0.2	0.5	0.8	28/97	9	9	10
0.2	0.7	1.0	24/84	13	6	5
0.6	0.3	0.4	28/92	13	7	8
0.6	0.6	0.9	28/93	11	8	9

Table 5：悪い例

λ	ρ	t		$\phi_k = 0.5$	$\phi_k = \widehat{\phi_k}$	$\phi_k = 0$
0.1	0.2	0.7	286/1130	8	139	139
0.2	0.2	0.9	254/1009	6	124	124
0.3	0.6	0.7	293/1159	10	141	142
0.3	0.9	0.4	299/1184	9	145	145
0.6	0.5	1.0	292/1150	15	138	139
0.6	0.8	0.5	252/991	14	119	119
0.7	0.2	1.0	259/1020	12	123	124
0.7	0.3	0.8	234/917	11	111	112
0.7	0.7	0.3	274/1083	9	132	133
0.8	0.8	0.7	287/1127	18	135	134

という結果が得られた。

Table 1 と Table 4 において一つだけ良いパラメータが共通しているが、他の数値実験データを見てみると必ずしも良いという結果は出ておらず、現時点では具体的に良いパラメータの値や範囲などを見つけることはできていない。今後はより多くの実験を通じて見つけていくことが課題となる。また、そのパラメータの選び方と関数の特徴に何らかの関係があるのかということも研究する必要がある。全体を通してよいパラメータが見つからなくても、何らかの特徴を持った関数に関してよいパラメータが発見できれば、それは今後の研究において有意義なものとなる。

それと同時に、Table 2 と Table 5 を見ればわかるように、ϕ_k の値として $\widehat{\phi_k}$ と 0 がほぼ同じ回数採用されている。この現象は、反復回数の多いパラメータの組み合わせすべてにおいて見て取れる現象であり、$\widehat{\phi_k}$ と 0 が交互に現れるということも確認されている。この現象は非常に興味深いものであり、どうしてそのような事実がおきているのかなどを調査するのが今後の課題である。

References

[1] Y.H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Applied Mathematics and Optimization* 43 (2001), pp.87-101.

[2] Y.H. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J.Optim.* 10 (1999), pp.177-182.

[3] J.J.Moré, B.S.Garbow and K.E.Hillstrom, Testing unconstrained optimization software, *ACM Transactions on Mathematical Software* 7 (1981), pp.17-41.

[4] A.Perry, A modified conjugate gradient algorithm, *Operations Research* 26 (1978), pp.1073-1078.

[5] H. Yabe and N. Sakaiwa, A New Nonlinear Conjugate Gradient Method for Uncostrained Optimization, *Technical Report, Department of Mathematical Information Science*, Tokyo University of Science, March, 2003.

[6] H. Yabe and M. Takano, Global Convergence Properties of nonlinear conjugate gradient methods with odified Secant condition, *to appear in Computational Optimization and Applications.*

[7] J.Z. Zhang, N.Y. Deng and L.H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, *J. Optim Theory Appl.* 102 (1999), pp.147-167.

[8] J.Z. Zhang and D.X. Xu, Propeties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, *J.Comput. Appl. Math.* 137 (2001), pp.269-278.